

# ECON832 Final: Mini Paper

Daniel Sánchez Pazmiño

April 2024

## 1 Overview

The data comes from the CPC18 (Plonsky, Erev, and Ert 2017), which consists of experiments involving various decision makers choosing between two lotteries. I focus on the newer data exclusive to CPC18 (Experiment 1), which added feedback and attention features to the CPC15 data.

This experiment involved sixty choice problems for 240 participants. Two sets of problems were implemented, where participants faced one set of thirty problems. Each problem was faced for 25 trials, where the first five trials did not provide feedback about forgone and obtained payoffs. A trial involved choosing between two options, where each option (A or B) involved an underlying lottery with a probability distribution (Plonsky et al. 2018). I use the calibration data set to train the DFNNs and the **individual track data** (Track II) to test the model.

## 2 Methodology

### 2.1 Data preparation

The calibration dataset was filtered to only include Experiment 1 data. The outcome variable was the binary indicator of the choice of B over A. Since the variable was already dichotomous, no transformation was needed. I include demographic variables in the baseline models: location, gender and age. For location and gender, I dichotomized the variables with Technion and male as the reference levels.

Risk features which are considered include the shapes of the distributions for lotteries A and B. These involve the number of lottery outcomes, the expected values, the low payoffs, and probabilities to draw from the lotteries and lot distribution shapes. The latter defines the distribution of the lotteries, which can be binomial around the expected value, right skewed or left skewed. The shape of the lotteries were dichotomized in the same way as the location and gender variables; all other variables are continuous. The variance, while not directly observed, is a function of these variables.

Other relevant variables involve whether there is ambiguity in the probabilities or if there is a correlation between the payoffs. Further, it is important to include trial numbers, potential and obtained payoffs obtained, and forgone payoffs. Some of these variables are categorical, but they are already dichotomized, so I didn't transform them.

Regarding attention features, I include the number of time block, reaction time, feedback binary outcome, the onscreen button and the trial number within a game.

The processing of the calibration dataset was done using `TidierData.jl` and `DataFrames.jl`. I performed a test-train split with 80% of the data used for training and 20% for testing. I further tested with the individual track data Plonsky, Erev, and Ert (n.d.).

## 2.2 Feature engineering

As mentioned, all categorical variables for both the risk and attention features were dichotomized. This is equivalent to **one-hot encoding**. The continuous variables were **standardized** to have a mean of zero and a standard deviation of one.

The data was processed from its rectangular form (every row per decision) to a format where each row represented an attribute, and every column represented a decision.

## 2.3 Model configuration

The model is a deep feedforward network neural network (DFNN) with `Flux.jl`. I use a ReLU activation function for the input and hidden layers and a sigmoid activation function for the output layer. The sigmoid function is commonly used for binary classification problems, as it squashes its input values to be between 0 and 1, which can be interpreted as probabilities.

I use mean squared error as the loss function and the ADAM optimizer to minimize the loss function. The learning rate was set to 0.5. I use a confusion matrix and an accuracy rate (percent correctly predicted) to evaluate the model's fit.

# 3 Feature analysis

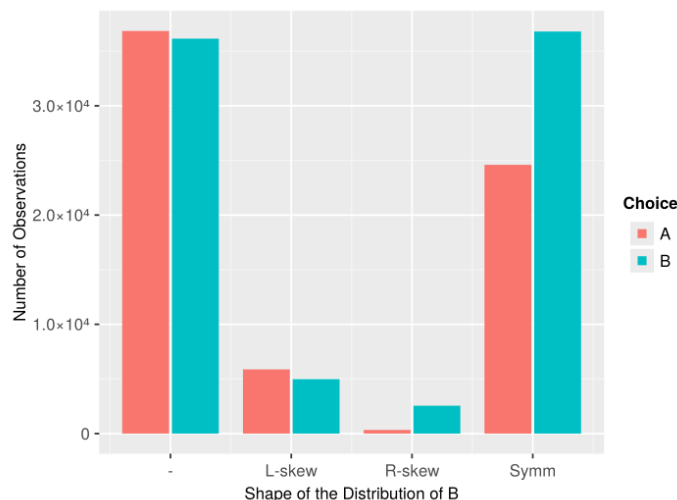
## 3.1 Risk features

A correlation analysis was performed to identify the most relevant features. The correlation matrix shows that the number of outcomes, the low payoff of lotteries, the correlation between two possible options and the shape of the distribution are the most relevant features.

The correlation between the number of outcomes and the choice for the B lottery is 0.09, which is statistically significant at the 1% level. This means that more outcomes in the B lottery are associated with a higher probability of choosing B. The correlation between the low payoff of the lotteries and the choice for the B lottery is 0.11, which is statistically significant at the 1% level. This means that higher low payoffs in the B lottery are associated with a higher probability of choosing B. This might be a result of the risk aversion of the participants, which is a common feature in decision-making and point towards the importance of considering risk-aversion in the model rather than risk neutrality (since the expected value is not correlated). The correlation between lotteries shows a relationship with the choice of B, with negative correlation of -0.14, which is statistically significant at the 1% level. This means that the correlation between the payoffs of the lotteries is associated with a lower probability of choosing B. Again, this might point toward risk aversion, as the correlation between payoffs might be seen as a riskier option.

In terms of the shape of the distribution, I identify that a symmetric distribution greatly favours the choice of B, as seen below. Further, I also identify that the payoff that the participant received shows a statistically significant correlation with the choice of B. I choose to include all 12 variables which jointly define the shape of the distribution in the model, as well as the payoff that the participant received plus the potential payoffs of A and B.

Figure 1: Choice of lottery by shape of distribution



## 3.2 Attention features

## References

- Plonsky, Ori, Reut Apel, Ido Erev, Eyal Ert, and Moshe Tennenholtz. 2018. “When and How Can Social Scientists Add Value to Data Scientists? A Choice Prediction Competition for Human Decision Making.” *Unpublished Manuscript*. <https://cpc-18.com/wp-content/uploads/2018/03/cpc18-white-paper-march-update.pdf>.
- Plonsky, Ori, Ido Erev, and Eyal Ert. 2017. “Calibration Data for Choice Prediction Competition 2018 (CPC18).” Data set with codebook. Zenodo. <https://doi.org/10.5281/zenodo.845873>.
- . n.d. “Competition Set Data Track II (Individual Behavior): Raw Data.” Data set. Zenodo. Accessed April 20, 2024. [CPC18: Choice Prediction Competition 2018](#).